Brooklyn Law Review

Volume 65
Issue 1
The Eighth Abraham L. Pomerantz Lecture

Article 6

9-1-1999

Causation, Mental Models, and the Law

P.N. Johnson-Laird

CAUSATION, MENTAL MODELS, AND THE LAW*

P.N. Johnson-Laird†

"[I]n all legal systems, liability to be punished or to make compensation frequently depends on whether actions (or omissions) have caused harm."[1]

INTRODUCTION

The accused assaulted and raped a woman, who then took poison and died either from the poison or its effects combined with those of her wounds. Did the accused cause her death? The defendant negligently left open an unguarded lift shaft, and a young lad, knowing that the lift was not there, invited the plaintiff to step into it. Did the defendant cause the plaintiff's injuries? The defendants, manufacturers of guns, negligently oversupplied shops in states with weak gun laws. Criminals came into possession of such guns and murdered relatives of the plaintiffs. Were the defendants the cause of the homicides?

Lawyers think about causal relations, because, as these examples show, liability can depend on causation. Lawyers also complain that people accept mere association as evidence

---

† Stuart Professor of Psychology, Princeton University; B.A. (Hons), 1964, University College London; Ph.D., 1967, University College London. I am grateful to the Center for the Study of Law, Language and Cognition for inviting me to give the inaugural lecture at the Center, and I am grateful to Larry Solan for inviting me to think about the topic. I also thank John Darley for his advice and expertise on bringing together psychological experiments and legal concepts, and my brother Andy Johnson-Laird for his many discussions over the years of legal problems and forensic software analysis. The theory of causation presented in the paper is a result of a collaboration with my graduate student, Yevgeniya Goldvarg, and I thank both her and many colleagues for their helpful advice: Victoria Bell, Ruth Byrne, Zachary Estes, Vittorio Girotto, Denis Hilton, Paolo Legrenzi, Bradley Monton, Hansjoerg Neth, Mary Newsome, David Over, Vladimir Sloutsky, Jean-Baptiste van der Henst, and Yingrui Yang.

[1] H.L.A. HART & TONY HONORÉ, CAUSATION IN THE LAW 63 (2d ed. 1985).

of causation.[2] Of course, more is at stake in law than causal relations, and in many cases they are irrelevant. Yet, in ordinary life, we blame people because they have caused harm to others; the law refines this notion and punishes individuals, or makes them liable for compensation, if they are responsible.

In pioneering studies, Solan[3] and Winter[4] have shown that cognitive and linguistic considerations can elucidate many problems in legal interpretation. The aim of the present paper is to clarify the concept of causation—the everyday notion that underlies common law. This paper also relies on the methods of cognitive psychology—experimentation and computer modeling of thought processes—to lay bare the underlying structure of causal relations in a way that is not normally available to legal theorists. Whether or not causation is important to the law, however, is a matter that the paper will not assess. It will pretend that those who seek to minimize the importance of cause in the law are wrong. Perhaps they are wrong; however, the pretence is warranted if the law should ultimately correspond to the conceptions of the community.[5]

The plan of the paper is simple. It begins with a sketch of a theory of how people understand and reason. This particular theory of cognition rests on the assumption that to grasp the meaning of a description is to envisage how the world would be if the description were true. The mind constructs *mental models* of the world, and each such model corresponds to a possible state of affairs. The paper then describes how people understand causal relations. People draw a distinction in meaning and in reasoning that is contrary both to recent probabilistic theories[6] and to a philosophical tradition going back at least to John Stuart Mill's analysis of causation.[7] The paper concludes

---

[2] *See, e.g.*, Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1090 (E.D. Md. 1986).

[3] *See* Lawrence Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt About Reasonable Doubt*, 78 TEX. L. REV. (forthcoming 1999).

[4] *See* STEVEN L. WINTER, A CLEARING HOUSE IN THE FOREST: HOW THE STUDY OF THE MIND CHANGES OUR UNDERSTANDING OF LIFE AND LAW (forthcoming 2000).

[5] *See generally* P.H. ROBINSON & J.M. DARLEY, JUSTICE LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW (1995).

[6] *See, e.g.*, P. SUPPES, PROBABILISTIC METAPHYSICS (1984).

[7] *See generally* JOHN STUART MILL, A SYSTEM OF LOGIC (J.M. Robson ed., University of Toronto Press 1973) (1843).

with an account of how the newly revealed structure of causation may elucidate some legal puzzles.

As Hume remarked, reasoning about matters of fact is largely based on causal relations.[8] Theorists in many disciplines have studied how humans think about causal relations, but they have yet to agree on a theory of that process. Scholars also disagree about the meaning of causal claims and their philosophical underpinnings. Our concern is not with philosophical problems—that is, with whether causal relations are objective states in the world or subjective notions in the mind, or with whether causal relations hold between facts, events, objects, or states of affairs. In fact, causes can concern events and also states of affairs, e.g., as Benjamin Franklin wrote, "for want of a nail the shoe was lost."[9] And so this paper uses the neutral expression, "states," to embrace physical or psychological events, situations, facts, and other potential arguments of causal relations. We pass over these fine philosophical distinctions, not because they are unimportant, but because their proper analysis would take us too far afield from this paper's main concerns.

I. THE THEORY OF MENTAL MODELS

How do human beings reason? One answer, which goes back to the Enlightenment, and which still has many adherents, is that humans follow the laws of thought.[10] These laws are made explicit in formal logic, the probability calculus, and the theory of rational decision making. Undoubtedly, people can acquire such laws and even use them to solve difficult problems. The question at issue, however, is whether *naive* individuals, i.e., those with no training in formal calculi, unconsciously follow the laws of thought. They certainly are not aware of following such laws. They cannot describe them. And several sorts of observation imply that they are not following

---

[8] *See* D. HUME, AN ENQUIRY CONCERNING HUMAN UNDERSTANDING (A. Flew ed., 1988).

[9] BENJAMIN FRANKLIN, POOR RICHARD'S ALMANACK (1758), *quoted in* JOHN BARTLETT, FAMILIAR QUOTATIONS 347 (Emily Morison Beck ed., 15th ed. 1980).

[10] *See, e.g.*, MENTAL LOGIC (M.D.S. Braine & D.P. O'Brien eds., 1998).

them. One such observation is that people make errors in reasoning. Here is an example based on Alan Dershowitz's presentation at the O.J. Simpson trial:

> If a man batters his partner, then he is unlikely to murder her. The FBI reports only about 1500 such cases out of four million cases of battery. Therefore, if a man murders his partner, then he is unlikely to have battered her. "Battery, as such, is not a good independent predictor of murder."[11]

The error will be explained later.

Sporadic errors in reasoning hardly refute the doctrine of the laws of thought, but what is embarrassing are systematic and predictable errors. The laws of thought, by definition, do not allow for them. The laws yield only valid inferences, so whatever errors occur should be sporadic and haphazard—a result of a contingent "accident in the mind" rather than basic principles. In fact, experiments carried out in the psychological laboratory show that people do make systematic and predictable mistakes in reasoning. Before discussion of such studies, we will first consider the theory of mental models, because it is this theory that predicted the errors.

The theory of mental models has its origins in a conjecture made more than fifty years ago by a prescient Scottish psychologist, the late Kenneth Craik.[12] He suggested that the mind builds small-scale models of the world, which it uses to anticipate events and to guide its decisions. Mental models are constructed as a result of perceiving the world, understanding descriptions, and imagining possibilities. This modern theory of mental models adds three important assumptions:[13]

> 1. Each mental model represents a possibility, and its structure corresponds to the structure of what it represents. For example, a model of a set of individuals, such as some lawyers in a court, consists in a set of mental tokens, where each token corresponds to an individual.

---

[11] ALAN M. DERSHOWITZ, REASONABLE DOUBTS: THE O.J. SIMPSON CASE AND THE CRIMINAL JUSTICE SYSTEM 104-05 (1996).

[12] *See* KENNETH CRAIK, THE NATURE OF EXPLANATION (1943).

[13] *See* P.N. JOHNSON-LAIRD, MENTAL MODELS: TOWARDS A COGNITIVE SCIENCE OF LANGUAGE, INFERENCE AND CONSCIOUSNESS (1983); P.N. JOHNSON-LAIRD & R.M.J. BYRNE, DEDUCTION (1991).

2. Mental models normally represent what is true, and not what is false. This postulate is so important that it is dignified with a name: the principle of *truth*.
3. A set of models may concern what is physically possible, what is permissible, or what is logically possible.

Reasoning calls for individuals to envisage the state of affairs corresponding to some starting point—a set of verbal premises, some piece of general knowledge, a visual observation, an assumption made for the sake of argument, or some admixture of them. To make matters simple, this paper refers henceforth to this starting point as "the premises," but readers should bear in mind that they may have a varied provenance. An inference is logically *valid* if its conclusion must be true given that its premises are true. While there is no guarantee that the conclusion of a valid inference is true, it will be true provided that its premises are true. If its premises are false, the conclusion may, or may not, be true.

The model theory provides a unified account of various sorts of reasoning. A conclusion is necessary—it *must* be the case—if it holds in all the models of the premises; it is possible—it *may* be the case—if it holds in at least one of the models; and its probability—assuming that each model is equiprobable—depends on the proportion of models in which it holds.

The theory can be illustrated by a simple inference. Consider the following problem:

The battery was either dead or else the starter did not work.
In fact, the battery was not dead.
What follows?

The first premise, which is an exclusive disjunction, is consistent with two possibilities:

dead
¬work

Each line in this diagram denotes a separate mental model corresponding to a separate possibility, and so "dead" denotes a model of the battery as dead. The symbol "¬" denotes negation, "work" denotes a model of the starter as working, and so "¬work" denotes a model of the starter as not working. Actual mental models can take the form of visual images or complex abstract structures, but we do not need to represent these de-

tails here. More importantly, the first model represents the truth of the proposition that *the battery was dead*, but it does not represent that it is false that *the starter did not work*, i.e., the starter did work. People try to make mental footnotes about what is false, but, as we will see, they soon forget them. The second premise eliminates the first model, and so the conclusion that follows corresponds to the second model:

The starter did not work.

This conclusion is valid, i.e., it is necessarily true given that the premises are true, and its validity is established by showing that it holds in all possible models of the premises. In this case, of course, there is only one model of both premises.

Suppose that we were to construct models that represent both what is true and what is false. The result, which our computer program works out for us, is a set of *fully explicit* models. They tell us what the correct conclusions are from a set of premises. In principle, people can construct fully explicit models if they can remember their mental footnotes about what is false. They can use the footnotes to flesh out mental models in order to turn them into fully explicit models. Hence, the mental models above of the exclusive disjunction can be fleshed out into fully explicit models:

dead	work
¬dead	¬work

All the theory's predictions about reasoning derive from the preceding account. The last twenty years has seen an accumulation of experimental evidence corroborating mental models, and over 250 papers have been published about them.[14] A major experimental result is that reasoners soon forget about what is false, especially with complex premises, and so they make egregious errors. Inferences that depend on a greater number of mental models are more difficult, taking longer and leading to more errors. Erroneous conclusions tend to correspond to individual mental models of premises. It is beyond the scope of this paper to review the evidence, but it is appropriate

[14] *See Mental Models in Reasoning* (visited Aug. 20, 1999) <http://www.tcd.ie/Psychology/Ruth_Byrne/mental_models> (webpage developed by Ruth Byrne and her colleagues at Trinity College, Dublin).

to show how the model theory explains errors in probabilistic reasoning, such as the error in the inference about battering and murder.

The model theory's representation of a conditional of the form:

If he battered his partner then he murdered her,

calls for models of the following form:

Batter Murder
...

where the ellipsis represents those possibilities in which the antecedent of the conditional, he battered his partner, is false. Thus, the ellipsis is really a "place holder" representing possibilities that people do not normally think about. But, if pushed, they may be able to make them explicit. Given that it is false that he battered his partner, he may or may not have murdered her, and so the fully explicit models of the conditional above are:

Batter Murder
¬Batter Murder
¬Batter ¬Murder

The *mental* models of the conditional resemble those of the converse conditional:

If he murdered his partner then he battered her,

which are of the form:

Murder Batter
...

The theory accordingly predicts that individuals will readily confuse the two assertions.

They are also likely to confuse analogous assertions about conditional probabilities. It may be true, as data from the FBI suggest:

If a man battered his partner, then he is unlikely to have murdered her (less than one case per 1000 cases of battery in 1992).

But, this claim is compatible with the *falsity* of the converse conditional:

If a man murdered his partner, then he is unlikely to have battered her.

With no further information, other than the assumption that there are, say, tens of thousands of couples who live together amicably (without battery or murder), we could fill in the set of fully explicit models in a variety of ways. One way conforms with the truth of the preceding conditional:

Frequencies		
Batter	Murder	1
Batter	¬Murder	999
¬Batter	Murder	99
¬Batter	¬Murder	44,901

This distribution of frequencies shows that if a man murdered his partner (the first and third rows), then he is most unlikely to have battered her (one case out of a 100 murders). But the data could equally well turn out to be as follows:

Batter	Murder	1
Batter	¬Murder	999
¬Batter	Murder	1
¬Batter	¬Murder	44,999

It is still true that:

> If a man battered his partner, then he is unlikely to have murdered her (one chance in a thousand).

But, the claim that:

> If a man murdered his partner, then he is unlikely to have battered her,

is utterly false. Indeed, the data show that given that a man murdered his partner (the first and third rows), there is a 50% chance that he battered her.

This cautionary tale has several morals. Reasoning about probabilities is difficult: we all make mistakes. None of us appears to be unconsciously equipped with the laws of the probability calculus. We often fail to realize that the conditional probability of one state A given another state B tells us nothing about the converse conditional probability of B given A. The secret for obtaining the right answer is to work out all the possibilities—to construct their fully explicit models—and to use the evidence at hand to fill in their frequencies of occurrence. If it is impossible to fill in the numbers completely, then we should beware of assertions about the probability of one state given another. But once we have filled in the fully explic-

it possibilities with their frequencies, we know all that we need to know in order to work out any probability concerning the states. An analogous method is also the secret to working out the correct causal relations among states.

II. THE MEANING OF CAUSAL RELATIONS

There are three main sorts of causal assertion:

1. *General* causal assertions, such as:

   Depriving a patient's brain of oxygen for five minutes causes death.

2. *Singular* causal assertions where the outcome is known, such as:

   Depriving this patient's brain of oxygen for five minutes caused his death.

3. *Singular* causal assertions where the outcome is not known, such as:

   Depriving this patient's brain of oxygen for five minutes will cause his death.

The first assumption of the model theory of causal relations is that they concern possibilities:

> Given two states of affairs, A and B, the meaning of a causal relation between them concerns what is possible and what is impossible in their co-occurrence.

The second assumption is that B does not precede A in time:

> Given two states of affairs, A and B, if A has a causal influence on B, then B does not precede A in time.

This principle allows, as Kant observed, that a cause can be contemporaneous with its effect.[15] For example, squeezing the toothpaste tube can occur at the same time that the toothpaste comes out of the tube. The meanings of causal relations do not call for action on contact because it is entirely proper to make causal claims relating distant states of affairs:

> The moon causes the tides.

---

[15] *See generally* IMMANUEL KANT, CRITIQUE OF PURE REASON (J.M.D. Meiklejohn trans., E.P. Dutton & Co. 1934) (1781).

The law is usually concerned with singular causal assertions, and so we will focus on them. A singular causal assertion where the outcome is not known, *A will cause B*, has three fully explicit models, which each represent a possibility:

1. a b
   ¬a b
   ¬a ¬b

These possibilities underlie the notion that A is *sufficient* for B to occur, that is, the causal relation is a weak one allowing for other causes of B. An example of this sort of weak causation is:

Increasing the amount he eats will cause him to gain weight,

because he can also gain weight, for example, as a result of exercising less. The claim, *A will cause B*, is false in case A occurs without B:

a ¬b

A singular causal assertion where the outcome is known, *A caused B*, has a model of the factual situation, and alternative models representing counterfactual possibilities:

a	b	(the factual case)
¬a	b	(a counterfactual possibility)
¬a	¬b	(a counterfactual possibility)

A counterfactual possibility is a state that was once possible but that did not, in fact, occur.[16] *A caused B* is false, of course, if either A or B did not occur. But, it can be false even when both A and B occurred, if the relation between them was not causal. Here the counterfactual possibilities are critical. If they include a case in which A occurred without B, then A did not cause B.

There are a variety of other causal relations, and each of them can occur in general and singular assertions. For simplicity, however, we consider only their occurrences as singular causal relations with unknown outcomes. The relation *A will*

[16] *See, e.g.*, R.M.J. Byrne, *Cognitive Processes in Counterfactual Thinking About What Might Have Been*, *in* 37 THE PSYCHOLOGY OF LEARNING AND MOTIVATION, ADVANCES IN RESEARCH AND THEORY 105-54 (D.K. Medin ed., 1997).

prevent B means that the occurrence of A will cause B not to occur. It has the following fully explicit models:

2. a ¬b
   ¬a b
   ¬a ¬b

An assertion of the form *A will allow B*, such as:

Taking the short cut will allow you to avoid the traffic,

has a strong *implicature* that not taking the short cut will not allow you to avoid the traffic. An implicature is an inference that is warranted by pragmatic considerations.[17] Hence, individuals who speak in an informative way would not utter the remark above if they knew that you could just as well avoid the traffic by not taking the short cut. Thus, the models of *A will allow B* and its implicature have the form:

3. a b
   a ¬b
   ¬a ¬b

These possibilities underlie the notion that A is *necessary* for B to occur. An assertion of the form:

A will not allow B,

and its implicature have the following models:

4. a ¬b
   a b
   ¬a b

It is sometimes important to distinguish between meaning and implicature, but in what follows we will not attempt to keep them apart.

The preceding relations are weak because they are consistent with three possibilities. In addition, however, there are two strong causal relations consistent with only two possibilities. *A and only A will cause B* has the following models:

5. a b
   ¬a ¬b

---

[17] *See* H.P. Grice, *Logic and Conversation*, *in* 3 SYNTAX AND SEMANTICS (P. Cole & J.L. Morgan eds., 1975).

And *A and only A will prevent B* has the following models:

6. a ¬b
   ¬a b

An example of strong causation is:

Too much alcohol will cause him to get drunk,

because drunkenness has no other cause.

The preceding analysis of the meanings of causal relations is in terms of fully explicit models. The principle of truth, however, predicts that naive individuals will tend to rely on the corresponding *mental* models for each of the causal relations. An assertion of the form:

A will cause B,

like the analogous assertion, *If A then B*, calls for the mental models:

a b
...

in which there is only an implicit model representing the possibilities in which the antecedent, A, is false. There is a mental footnote to capture this information, and given the footnote, it is possible to flesh out the models fully explicitly as:

a b
¬a b
¬a ¬b

where B occurs in the absence of A.

The theory postulates that individuals normally reason on the basis of mental models, but with simple assertions of the present sort, they can appreciate that *A will cause B* is compatible with B having other causes. Given time, they may even enumerate the explicit possibilities. Strong causation as expressed by *A and only A will cause B* has the same mental models as weak causation, but the mental footnote indicates that the only way to flesh out the models fully explicitly is as:

a b
¬a ¬b

Table 1 summarizes the mental models of the six singular causal relations together with their fully explicit models. The same models stand for general causal assertions, but each row

then represents an alternative state within the same situation. The existence of several sorts of causal relation may come as a surprise to the reader. Philosophers have often assumed that there is only a single relation of cause and effect—an oversight that according to Hesslow[18] has led them into difficulties.

Table 1 — The models for six singular causal relations. The central column shows the mental models normally used by human reasoners, and the right-hand column shows the fully explicit models, which represent the false components of the true cases using negations that are true: "¬" denotes negation and ". . ." denotes a wholly implicit model. The mental models for the strong and weak relations of cause and prevention differ only in their mental footnotes (see text).

Connective	Mental Models	Fully Explicit Models
1. A will cause B:	A B . . .	A B ¬A B ¬A ¬B
2. A will prevent B:	A ¬B . . .	A ¬B ¬A B ¬A ¬B
3. A will allow B:	A B . . .	A B A ¬B ¬A ¬B
4. A will allow not B:	A ¬B . . .	A ¬B A B ¬A B
5. A and only A will cause B:	A B . . .	A B ¬A ¬B
6. A and only A will prevent B:	A ¬B . . .	A ¬B ¬A B

There is no causal relation if A or B is bound to occur, or bound not to occur, and experiments show that naive individu-

[18] *See* G. Hesslow, *The Problem of Causal Selection, in* CONTEMPORARY SCIENCE AND NATURAL EXPLANATION: COMMONSENSE CONCEPTIONS OF CAUSALITY 11-32 (D.J. Hilton ed., 1988).

als concur with this claim. If all you know is that B is bound to occur if A occurs, then you are not entitled to make any causal claim about the relation between them. On the one hand, B may also be bound to occur even if A does not occur, i.e., there are only the following possibilities:

a b
¬a b

Thus, it would be wrong to invoke any causal relation between A and B in this case. On the other hand, even if there is a causal relation between the two, you do not know whether A is a strong or weak cause of B. Yevgeniya Goldvarg and I have carried out experiments in which naive individuals had to list what was possible and what was impossible given various causal relations. The results corroborated the model theory. No participant balked at the task, and most of them were able to generate fully explicit models, distinguishing among the various causal relations, and between causing and allowing.

III. Deductive Inferences from Causal Relations

Some theorists suppose that causes can be neither observed nor deduced, and so they are only induced. But the causal status of an observation *can* be deduced from a knowledge of its circumstances. For example, suppose you know that a certain anesthetic causes a loss of consciousness, and you observe that a patient who is given the anesthetic loses consciousness. You can deduce that the anesthetic *caused* the patient to lose consciousness.

How do you make causal deductions? One answer is that you rely on the laws of thought, that is, formal rules of inference of some sort. Another answer, however, is that you rely

on mental models. Given the appropriate temporal constraints and the following set of possibilities:

a	b	c
¬a	¬b	c
¬a	b	¬c

you validly infer:

A will cause C,

because the set contains each of the possibilities required for this relation. In practice, as Goldvarg and I have shown experimentally, naive individuals tend to rely, not on fully explicit models, but on mental models.[19] For example, given a problem of the form:

A prevents B.
B causes C.
What, if anything, follows?

Most people drew the conclusion:

A prevents C.

The mental models of the premises support this conclusion, but the fully explicit models show that it is wrong: there is no causal relation between A and C. In general, when naive individuals make a causal inference, they tend to err when there is a discrepancy between the conclusion supported by the mental models of the premises and the conclusion supported by the fully explicit models of the premises. They tend to draw the conclusion corresponding to the mental models.

The most striking result depends on an unexpected prediction. Because mental models fail to represent what is false, they yield grossly erroneous conclusions from certain premises. These premises give rise to *illusory* inferences, i.e., most people draw one and the same conclusion, which seems obvious, and yet which is wrong. Such illusory inferences occur in other domains of reasoning. These illusory inferences provide a

[19] *See* Table 1.

strong support for the model theory, because they are contrary to theories of reasoning based on the "laws of thought." As an example, consider the following problem:

> One of these assertions is true and one of them is false:
> Marrying Pat will cause Viv to relax.
> Not marrying Pat will cause Viv to relax.
> The following assertion is definitely true:
> Viv will marry Pat.
> Will Viv relax?

The mental models of the disjunction of the first two premises are as follows:

Marry	Relax
¬Marry	Relax

and so it seems that Viv is bound to relax. But, these models fail to represent what is false, i.e., when the first premise is true, the second premise is false, and *vice versa*. For example, if it is false that marrying Pat will cause Viv to relax, then Viv will not relax even though Viv marries Pat:

Marry	¬Relax

Hence, the premises do not imply that Viv will relax. The conclusion is an illusion. Nearly everyone succumbs to illusory inferences, and thereby corroborates the model theory.

IV. HART AND HONORÉ'S THEORY OF CAUSATION

Hart and Honoré's classic book, CAUSATION IN THE LAW,[20] which was originally published in 1959, describes a theory of causation that is incompatible with the previous account based on models of possibilities. Many theorists have argued that some causal power or mechanism over and above what can be captured in possibilities somehow links causes to effects. Likewise, a central component of Hart and Honoré's analysis is that singular causal assertions rely implicitly on general causal claims. They write:

> even singular causal statements which appear to be confined to the connection between two particular occurrences are in fact covertly general; their causal character is derivative and lies *wholly* in the

[20] *See* HART & HONORÉ, *supra* note 1.

> fact that the particular events with which they are concerned exemplify some generalization asserting that kinds or classes of events are invariably connected.[21]

Certainly, such generalizations often exist, and they can play a crucial role when individuals induce a causal relation from observations or assess whether a singular causal assertion is true. But they are not part of the *meaning* of causation. Assertions that explicitly deny the existence of general principles are not self-contradictory, e.g.:

> Psychokinesis caused the bias in the random-number generator, though it did not exemplify any general relation and how it had such an effect is in principle inexplicable.

The assertion may be false, but it is not a self-contradiction, as it would be if the meaning of *causes* entailed a background generalization.

Hart and Honoré do not explicitly distinguish between causing and allowing (i.e. so-called "enabling conditions"), but they distinguish between causes and what they refer to as "mere conditions."[22] Their argument, which is similar to Mill's,[23] can be illustrated by an example of a singular cause.

Consider an explosion that is caused by the occurrence of a spark in a container of combustible vapor. The explosion would not have occurred in the absence of the spark, and it would not have occurred in the absence of the vapor. Hence, both the spark and the vapor appear to be individually necessary and jointly sufficient to cause the explosion. Yet people normally speak of the spark as the *cause* of the explosion and the presence of the vapor as the *enabling* condition that allows it to occur. Likewise, the absence of the vapor would be a *disabling* condition that would not allow the explosion to occur. If the difference between causes and enabling conditions cannot be accounted for in terms of logic or meaning, then what distinguishes the two? Mill himself thought that in everyday life, the choice of a cause was often capricious, but he did offer some more systematic answers, as have many authors. Thus, Hart and Honoré argue that when individuals identify single causes, they choose the unusual or abnormal factor as the cause—the

---

[21] *Id.* at 10.
[22] *Id.* at 12.
[23] *See generally* MILL, *supra* note 7.

spark rather than the vapor in the example above or else they choose a voluntary human action.[24] There are many other schools of thought on this issue. Some say that the cause is the exceptional event, some say it is inconstant whereas the enabling conditions are constant, some say the cause is the factor that is relevant in explanations, and so on.

All these accounts could be true, but in contrast to them, the model theory draws a distinction in meaning and logic between causes and enabling conditions. Hence, this point is a crux for the model theory to which we will return in the next section. Meanwhile, we address a more general doubt about the model theory: is there really nothing more to the meaning of causal relations than possibilities and the temporal constraint that an effect does not precede its cause?

One doubt is that the mere existence of the relevant set of possibilities satisfying the temporal constraint does not suffice for a causal relation in certain instances. Granted the following two alternative possibilities, for example:

Divisible by two	Even
Not divisible by two	Not even

you do not assert:

> Being divisible by two causes a number to be even,

but rather:

> Being divisible by two *necessarily implies* that a number is even.

The domain is one of logical possibilities rather than physical possibilities. Likewise, the following alternative possibilities exist in a New Jersey bar:

Over the age of 18	Drinks alcohol
Over the age of 18	Does not drink alcohol
Not over the age of 18	Does not drink alcohol

Yet you do not assert:

> Being over the age of 18 enables a customer to drink alcohol,

but rather:

> Being over the age of 18 makes it *permissible* for a customer to drink alcohol.

---

[24] *See* HART & HONORÉ, *supra* note 1, at 34-35.

The domain is one of *deontic* possibilities rather than physical possibilities, that is, someone may, in fact, drink alcohol under the age of 18, but it would violate what is permissible. Causal, logical, and deontic claims all relate to possibilities, but the domain within which they hold (physical vs. inferential vs. deontic) yields the difference between them.

Yet, even in domains concerning physical possibilities, could there be cases that satisfy the model theory but that are not causal? Consider, for example, the relation between day and night: it is not possible to have day without night following closely on its heels. Or so it seems. Yet, day does not *cause* night. Hence, surely something is wrong with our analysis, and more is at stake than mere possibilities. Echoing Hart and Honoré's claim for background generalizations, theorists argue that what is missing is an *explanation* of the circumstances—day is a consequence of the sun shining on the earth, and, as the earth rotates, so day in a part of the earth gives way to night. Thus, the correlation between day and night is explained in terms of a common underlying cause. A corollary of this account is that we *can* envisage a possibility in which day is not followed by night: you need only orbit the earth at a speed that counteracts the effects of the earth's rotation, and you will live in perpetual day. This possibility refutes the causal claim that day causes night. We conclude that the invocation of explanatory principles and generalizations is a useful way—perhaps the only way in many cases—to infer causation from correlation, but the resulting conclusion of a causal relation means no more than that a certain set of temporally-ordered possibilities obtains.

V. CAUSE, ENABLING CONDITIONS, AND CIRCUMSTANCE

Suppose a witness testifies that the following sequence of events occurred:

> The doctor injected a drug into the patient and the patient lost consciousness.

What is the causal relation, if any, between the two events? The observation is inconsistent with two of the six possible causal relations: the injection did not prevent loss of consciousness in either the strong or weak sense. But, it is compatible with any of the four remaining relations and, of course, with the lack of any causal relation. A corollary of this uncertainty is that the mere observation of a particular sequence of states never suffices to establish a unique causal relation between them. Causal relations are modal. They are not merely about what occurred but also about what might have occurred. What might have occurred, however, cannot be determined merely from observation. It depends on a knowledge of the circumstances, i.e., the set of possibilities. The model theory accordingly postulates:

> Causal interpretation depends on how people conceive the *circumstances* of states, that is, on the particular states that they consider to be possible, whether real, hypothetical, or counterfactual.

Hart and Honoré had an analogous idea in mind when they talked of the "context" of a cause,[25] and other theorists have invoked similar ideas. The circumstantial principle, however, implies that individuals use their general knowledge and their knowledge of the state of affairs at issue to generate a set of mental models. Each model represents a possibility, and, in the case of a singular causal claim about a fact, one model represents the actual state of affairs. The models represent what a person takes to be the relevant possibilities in the circumstances, and they determine what the person judges to be the appropriate causal relation.

---

[25] *Id.* at 35.

To return to the example, if the circumstances are as follows:

injection	loss-of-consciousness
¬injection	loss-of-consciousness
¬injection	¬loss-of-consciousness

an appropriate description is:

The injection caused the patient to lose consciousness.

If the circumstances are as follows:

injection	loss-of-consciousness
injection	¬loss-of-consciousness
¬injection	¬loss-of-consciousness

an appropriate description is:

The injection allowed the patient to lose consciousness.

And if the circumstances are as follows:

injection	loss-of-consciousness
injection	¬loss-of-consciousness
¬injection	loss-of-consciousness

an appropriate description is:

The injection did not prevent the patient from losing consciousness.

What are the correct circumstances of a state of affairs? There may be no definite way to decide. That is why they bedevil both legal and philosophical analyses of causation. For example, inferences of the form known as "strengthening the antecedent" are valid in many circumstances:

A causes C.
∴A and B cause C.

This inference, for example, is unexceptional:

Putting sugar in your tea causes it to taste sweet.
∴Putting sugar in your tea and putting some milk in too causes it to taste sweet.

In contrast, the following inference is unacceptable:

Putting sugar in your tea causes it to taste sweet.
∴Putting sugar in your tea and putting some diesel oil in too causes it to taste sweet.

The circumstances of the conclusion are no longer those of the premises. The case is altered; the conclusion is false even if the premise is true.

The model theory distinguishes between causing an effect and allowing it to occur, that is, between causes and enabling conditions. Yet, theorists such as Mill and Hart and Honoré deny that there is any logical or semantic distinction between the two. Hence, an apparent paradox needs to be resolved. We will show that the circumstances of states do indeed resolve the controversy: causing *is* logically distinct from allowing. And we will spell out why Mill's argument seemed compelling to Hart, Honoré, and to their peers.

We argued earlier that there is a difference in meaning between *A will cause B* and *A will allow B*. The difference is borne out by their logical consequences. For example, suppose that the following proposition is true:

> Taking bichloride of mercury will cause her to die,

and that the causal antecedent is true:

> She takes bichloride of mercury.

You can draw the valid conclusion:

> She will die.

In contrast, suppose that the following proposition is true:

> Not taking bichloride of mercury will allow her to live,

and that the enabling antecedent is true:

> She does not take bichloride of mercury.

You can draw only the weak conclusion:

> She may live.

It follows that the two sorts of relation *are* logically distinct.

Granted the distinction, why should anyone ever have assumed that causes and enabling conditions are logically indistinguishable? The answer comes from the subtle effects of knowledge in determining the circumstances of an effect. Consider, again, the example that we used to illustrate Mill's argument. A spark in a combustible vapor causes an explosion. You know that in the presence of the vapor, the spark causes the

explosion, and that in the absence of either the vapor or the spark, there is no explosion. Your knowledge accordingly yields the circumstances shown in the following models:

vapor	spark	explosion
vapor	¬spark	¬explosion
¬vapor	spark	¬explosion
¬vapor	¬spark	¬explosion

In these circumstances, the roles of the spark and vapor *are* equivalent. Jointly, they are the strong cause of the explosion.

You can envisage other circumstances, however. Suppose, for example, that a tank is used to store gasoline, and at present it may or may not contain a combustible vapor. If it does contain the vapor then the spark (or, say, a naked flame) will cause an explosion. The circumstances are shown in the following models:

vapor	spark	explosion
vapor	¬spark	explosion
vapor	¬spark	¬explosion
¬vapor	spark	¬explosion
¬vapor	¬spark	¬explosion

The respective probabilities of each of these possibilities can make one antecedent normal and the other antecedent abnormal, but they have no bearing on their causal roles.[26] What matters is that their respective roles are logically distinct in the circumstances. In these models, the relation between the vapor and the explosion is as follows:

vapor	explosion
vapor	¬explosion
¬vapor	¬explosion

Hence, the vapor allows the explosion to occur. In contrast, the spark and the explosion occur in all possibilities in the models above:

spark	explosion
spark	¬explosion
¬spark	explosion
¬spark	¬explosion

[26] *But see* HART & HONORÉ, *supra* note 1.

There is no causal relation here between spark and explosion, but the original set of models of the circumstances shows that the presence of the vapor enables the spark to cause the explosion, i.e.:

> Given the presence of the vapor, the spark causes the explosion, but in the absence of the vapor, there is not an explosion whether or not the spark occurs.

You can envisage still other circumstances in which the causal roles of the vapor and spark are interchanged. Suppose, for example, that an induction coil delivers a spark from time to time in an enclosed canister. You know that the introduction of a combustible vapor will cause an explosion. You also know that the occurrence of the spark allows such an explosion to occur. It may even occur without the vapor if, say, an explosive substance such as gunpowder is put into the canister. The circumstances are as follows:

spark	vapor	explosion
spark	¬vapor	explosion
spark	¬vapor	¬explosion
¬spark	vapor	¬explosion
¬spark	¬vapor	¬explosion

A description of these circumstances is:

> Given the occurrence of the spark, the vapor causes the explosion, but in the absence of the spark, there is not an explosion whether or not there is the vapor.

And, once again, this description is not affected by the rareness of the spark or the vapor.

This analysis shows that causes and enabling conditions are distinct, that they reflect the possibilities in the circumstances, and that they are not distinguishable merely on the grounds of relative rareness. Causes need not be unusual or abnormal, and they need not be pragmatically relevant to explanations. It is true that enabling conditions are constant in some circumstances. But, our preceding examples show that constancy in the circumstances is not necessary for an enabling condition. Neither the spark nor the vapor is constant in the circumstances above, yet their logical roles are distinct, and one is the enabling condition and the other the cause. What we can say, however, is that the cause is only effective given the presence of the enabling condition. Yevgeniya Goldvarg and I

have shown experimentally that naive individuals distinguish between causes and enabling conditions in scenarios that parallel our examples of sparks, vapors, and explosions.

VI. PROBABILISTIC THEORIES OF CAUSATION

Certain philosophers have proposed that causation is a probabilistic concept. Reichenbach[27] and others have defended such a notion, arguing that *A causes B* if:

$$p(B \mid A) > p(B \mid \neg A)$$

where $p(B \mid A)$ denotes the conditional probability of B given that A occurs, and $p(B \mid \neg A)$ denotes the conditional probability of B given that A does *not* occur. The probabilities can be computed from a distribution of the relative frequencies of the various possibilities:

A	B	28
A	¬B	3
¬A	B	10
¬A	¬B	59

The difference between $p(B \mid A)$ and $p(B \mid \neg A)$ is markedly positive (.90 - .14 = .76), and so according to Cheng, *A causes B*.[28]

The main evidence for a probabilistic semantics is that people judge that a causal relation holds in cases, such as the preceding example, in which the antecedent is neither necessary nor sufficient to bring about the effect. Most people, for instance, will assent to the proposition:

Smoking causes lung cancer,

even though they know that not everyone who smokes gets the disease. Such loose generalizations are common in daily life.

---

[27] *See* R. REICHENBACH, THE DIRECTION OF TIME (1956).

[28] *See* P.W. Cheng, *From Covariation to Causation: A Causal Power Theory*, 104 PSYCHOL. REV. 367 (1997).

Yet most people are also likely to assent to the more accurate proposition:

Smoking often causes lung cancer.

Readers who agree that this assertion is more accurate have conceded the main point: if causes were intrinsically probabilistic, then the two assertions would not differ in accuracy.

The probabilistic approach may be justified for scientific conceptions of causation, especially since the development of quantum mechanics. But it is implausible as an account of the meaning of causal relations in everyday life. Our main evidence against a probabilistic meaning for causality is that naive individuals tend to divide cases into those that are possible and those that are impossible given that *A causes B*. This result is contrary to a probabilistic theory, which allows that all cases are possible: it is their probabilities that matter. Likewise, consider the following equal distribution of frequencies:

Sunlight	Fertilizer	Growth	20
Sunlight	¬Fertilizer	Growth	20
Sunlight	¬Fertilizer	¬Growth	20
¬Sunlight	Fertilizer	¬Growth	20
¬Sunlight	¬Fertilizer	¬Growth	20

It follows that:

p(Growth | Sunlight) > p(Growth | ¬Sunlight), i.e., .66 > 0

and:

p(Growth | Fertilizer) > p(Growth | ¬Fertilizer), i.e., .5 > .33

Hence, both sunlight and fertilizer are causes of growth according to the probabilistic account, and sunlight is the stronger candidate. Yet, as the model theory predicts, individuals judge sunlight to be the enabling condition and fertilizer to be the cause. In short, the probabilistic theory obliterates the distinction between causes and enabling conditions.

Why do people so commonly assent to loose causal generalizations? One factor may be that they are aware that many causes in everyday life yield their effects only if the required

but unknown enabling conditions are present and the potentially disabling conditions are absent (as in the fertilizer example). It follows that when people assent to loose generalizations such as:

Smoking causes cancer,

they are granting that the causal relation holds unless some enabling condition is absent or some disabling condition is present.

VII. CAUSAL CHAINS AND CAUSAL FORKS

The great biologist and statistician Sir Ronald Fisher argued on behalf of the Imperial Tobacco Company that smoking might not be a cause of lung cancer.[29] He suggested instead that there could be an unknown gene, X, that both causes you to smoke and independently causes you to get lung cancer. If you have the deadly gene, then you are likely to develop the cancer whether or not you smoke. If you do not have the deadly gene, then you are unlikely to develop the cancer whether or not you smoke. Ergo, you might as well smoke. A contrasting view, of course, is that insofar as genes enter the picture there is a causal chain: the gene causes smoking, and smoking causes cancer. Thus, Fisher defended a causal fork of the form:

Gene —>Smoking
\—>Cancer

and his opponents defended a causal chain of the form:

Gene —>Smoking—>Cancer

where the arrows denote causation. How can we decide between these two contrasting accounts?

The computer program implementing the model theory shows us the differences in the two sets of possibilities corresponding to the chain and to the fork. Thus, given the assertions for the chain of weak causes:

Gene causes smoking.
Smoking causes cancer.

---

[29] *See generally* RONALD AYLMER FISHER, SMOKING, THE CANCER CONTROVERSY; SOME ATTEMPTS TO ASSESS THE EVIDENCE (1959).

the program constructs the following set of possibilities:

Gene	Smoking	Cancer
¬Gene	Smoking	Cancer
¬Gene	¬Smoking	Cancer
¬Gene	¬Smoking	¬Cancer

In contrast, given the assertions for the causal fork:

Gene causes smoking.
Gene causes cancer.

the program constructs the following possibilities:

Gene	Smoking	Cancer
¬Gene	Smoking	Cancer
¬Gene	Smoking	¬Cancer
¬Gene	¬Smoking	Cancer
¬Gene	¬Smoking	¬Cancer

The distinction is that the causal fork allows a possibility in which smoking occurs without cancer. Hence, there is crucial, though wholly unethical, experiment to decide between the two possibilities. A random sample of children is divided at random into two groups in order to ensure a roughly equal distribution of the gene in the two groups. One group is then forced to smoke, and the other group is prevented from smoking. The relative frequencies with which the two groups develop cancer will determine whether there is a fork or a chain. In Britain, the decisive evidence came from a study in which doctors gave up smoking, and it was possible to compare the rate at which they contracted cancer with the rate in a comparable group that continued to smoke. Even Fisher did not argue that there was a gene for giving up smoking in the cause of science.

Bradley Monton has pointed out an interesting set of circumstances. Consider these possibilities, which are presented in their temporal order:

First state	Second state	Third state
¬First state	¬Second state	Third state
¬First state	¬Second state	¬Third state

What is their correct causal description? According to the model theory, these circumstances correspond to both a fork:

> The first state is a strong cause of the second state, and a weak cause of the third state.

and to a chain:

> The first state is a strong cause of the second state, and the second state is a weak cause of the third state.

Surprisingly, both descriptions are correct. Hence, temporal constraints aside, if a first state is the strong cause of a second state, then any other state that is necessary given the second state, will be indistinguishable from a state caused by the first state. We are unable in principle to draw a distinction between this chain and fork, because it is impossible to have a case in which the second state does not yield the third state. There are, therefore, causal descriptions that seem to differ in meaning but that, in fact, are synonymous. They trade on the subtleties of strong causation, which has unforeseen and counterintuitive consequences.

VIII. CAUSATION AND THE LAW

This paper has described a new theory of causality as it is conceived by people with no training in philosophy, logic, or law. The theory provides accounts of what causal relations mean, of how they are mentally represented, and of how people make deductive inferences from them. It draws a sharp distinction between the meaning of causal relations and the evidence that supports them; and it distinguishes between causing an effect and allowing it to occur, not in terms of their normality, but in terms of their meanings. It also distinguishes between weak causation in which a state is sufficient, but not necessary, to bring about an effect, and strong causation in which a state is necessary and sufficient to bring about an effect. If this theory is correct, then many other theories are wrong. The principal consequences of the theory are that the meanings of causal relations are determinate, not probabilistic, and that they concern solely a temporally ordered set of possibilities. The assertion *A caused B* can accordingly be paraphrased as *A made it impossible for B not to occur*. A corollary is that *cause* is a transitive relation. The assertion that *A al-*

lowed B can be similarly paraphrased as *A made it possible for B to occur*. The case for a causal relation depends on observation, background knowledge, and common sense. None of these components, however, concern the meaning of the relation, but merely help us to determine whether or not that meaning is satisfied by states of affairs. The best test to establish a general causal relation is a scientific experiment. The best evidence for a singular causal relation is to show that it is an instance of a general causal relation.

The chief feature of the theory for legal matters is the profound difference it draws between the meaning of causal and enabling relations—a distinction that is robust in the performance of the lay individuals who participated in our experiments. In the past, legal theorists have tended to follow the line that the distinction is slightly capricious—to use John Stuart Mill's term[30]—and it accordingly appears to be often overlooked or blurred in legal judgments. Likewise, as Greene and Darley have shown,[31] the Model Penal Code appears to be based on assumptions about causation that are at odds with those of lay individuals. In particular, the Model Penal Code specifies that to be guilty of murder the accused's actions must be both the factual and the legal cause of the victim's death. To be the *factual* cause, the actions must have been necessary for the death, which implies that they must be a strong cause:

Actions	Death
¬Actions	¬Death

Legal cause, however, is a puzzling notion. The actions must lead to the death in a reasonably direct manner so that the causal relation is not too remote or indirect. Greene and Darley asked the participants in an experiment to make ratings of a set of scenarios about a murder attempt. The scenarios had a common beginning but divergent endings in terms of, for example, whether the murderous action led to the victim's death. In fact, the perpetrator's contribution to the outcome was the strongest predictor of the participants' ratings of liability. In contrast to the Model Penal Code, the

---

[30] *See* MILL, *supra* note 7, at bk. III, ch. v, § 3.

[31] *See* E.J. Greene & J.M. Darley, *Effects of Necessary, Sufficient, and Indirect Causation on Judgments of Criminal Liability*, 22 L. & HUM. BEHAV. 429 (1998).

participants treated actions that were weak causes, i.e., sufficient to bring about the victim's death, as calling for the severest punishment.

Another problematic aspect of causation in law occurs when courts are asked to determine not whether one state caused another, but rather whether one individual would have acted in a different way granted access to certain information. Thus, Twerski and Cohen[32] have argued that recent decisions in the law of informed consent have placed courts in an unworkable situation. Courts have to establish a causal connection between a doctor's failure to disclose information about a medical procedure and the patient's decision to undergo the procedure. The critical relation is whether, if the doctor had disclosed the risks, the patient would have declined the procedure (decision causation). The courts also have to establish a causal connection between the procedure and the patient's subsequent harm (injury causation). The difficulty, as these authors point out, is to establish what the patient would have done had he or she known the risks. Psychologists have shown that the mental processes of decision making are sufficiently complex that theorists can at best make only actuarial predictions, not predictions about the hypothetical counterfactual decisions of particular individuals.

The introduction presented three causal vignettes[33] that we will re-examine in order to illustrate how the model theory can elucidate some puzzles of causation. Consider the recent case against gun manufacturers:

> The defendants, manufacturers of guns, negligently oversupplied shops in states with weak gun laws. Criminals came into possession of such guns and murdered relatives of the plaintiffs. Some of the defendants were found to be the 'proximate' cause of the homicides.

The expression "proximate cause" is used in common law and in some statutes to bring out the point that it does not suffice merely to establish that one state is a necessary cause of another. What is meant by "proximate" is as puzzling as the concept of "legal" cause: perhaps the two denote the same concept. "Proximate cause" evidently does not refer to proximity in

---

[32] *See* Aaron D. Twerski & Neil B. Cohen, *Informed Decision Making and the Law of Torts: The Myth of Justiciable Causation*, 1988 U. ILL. L. REV. 607 (1988).

[33] *See supra* Introduction.

space or time, or to the most immediate cause in a chain of causes, though legal theorists have made both these interpretations.[34] Moreover, according to the model theory, the defendants were in no way a *cause* of a homicide. But they may have made it possible, that is, their negligence may have been an enabling condition. To prove this contention, you need to show that if they had not oversupplied guns to shops in states with weak gun laws, the guns would not have come into the possession of those criminals who committed the homicides. That is, you need, at the very least, to show that a homicide was committed using a gun that the defendants negligently supplied to a gun shop. Whether this case was established is unclear. Professor David Yassky was quoted in the *Guardian Weekly* to the following effect: "They split the baby. The jurors found that the industry was negligent in the way it sells and distributes guns, but they could not find a clear, strong link between that negligence and specific crimes."[35] A more general concern is whether the law of tort makes a sharp distinction between negligence that causes harm and negligence that enables harm to occur. Indeed, the court's decision in the present case contrasts starkly with the next case:

> The defendant negligently left open an unguarded lift shaft, and a young lad, knowing that the lift was not there, invited the plaintiff to step into it. Did the defendant cause the plaintiff's injuries?

In this case, the court followed the traditional view that a causal connection is negatived by a free action of a third party who exploits the situation created by the defendant. In fact, the defendant's negligent action merely enabled the accident to occur. Hence, the situation is analogous to the gun manufacturer's negligence. The shopkeepers exploited a situation created by the defendants and sold the guns to the criminals. The actions of the shopkeepers therefore might have been taken to negative the causal connection.

> The accused assaulted and raped a woman, who then took poison. The accused refused to summon medical help and imprisoned her for

---

[34] *See* HART & HONORÉ, *supra* note 1, at 86.

[35] Robert Suro, *Gun Makers Guilty of Negligence*, GUARDIAN WKLY., Feb. 21, 1999, at 15 (quoting Prof. David Yassky).

> several hours in a hotel room. She died a month later either from the poison or its effects combined with those of her wounds. Did the accused cause her death?

This case[36] presents a riddle to causal theorists. One reaction to the case is, in effect, why should we worry about the precise causal details of the victim's death when the court has before it a vicious criminal? Another reaction is that as the viciousness of intent increases so courts should be prepared to follow the consequences of an act farther and farther.[37] In fact, the Indiana court affirmed the conviction by a majority of three to two. The contrasting principle is, again, that if a free and voluntary act by another actor, who is not in concert with the first, intervenes and leads to harm, the first actor is relieved of responsibility. An appeal on this basis was rejected on the grounds that even if the victim had freely and voluntarily taken the poison, her wounds inflicted by the accused "actively contributed" to her death.[38]

The reasoning of the court in this case is odd. There are several causal questions to be answered. First, what was the cause of death: the wounds alone, the poison alone, or the two together? The court evidently decided that it was the two together. Second, did the behavior of the accused compel the victim to take the poison, that is, did she have sufficient reason to do so as a consequence of his actions or threats? If she was compelled, then the accused's actions caused her death. Third, if she was not compelled to take the poison, did the accused in preventing access to medical aid cause the victim to deteriorate in a way that was sufficient to lead to her subsequent death? These questions contain some subtleties, as the

---

[36] Stephenson v. State, 179 N.E. 633 (Ind. 1932).

[37] *See* HART & HONORÉ, *supra* note 1, at 99-100 (relevant citations).

[38] *See id.*

computer program implementing the model theory demonstrates. Given an input equivalent to the following propositions:

> The wounds (and threats) caused fear.
> Fear caused the victim to take poison.
> The poison and delay in medical aid caused death.

the program constructed the following set of possibilities:

Wounds	fear	poison	delay	death
Wounds	fear	poison	¬delay	death
Wounds	fear	poison	¬delay	¬death

and ten more possibilities in which there were no wounds or no fear. It follows that if the accused caused the wounds and the delay, they suffice for the victim's death (see the first possibility). Likewise, it is not necessary to show that the medical aid would have *caused* the victim to recover. That demonstration would be a mistake unless you also showed that it was a strong cause of recovery. It suffices to show merely that the aid might have enabled the patient to recover, i.e., with the aid she might or might not have recovered, but without it, she was bound to die. Given an input equivalent to the following propositions:

> The accused's action prevented medical aid.
> The medical aid allowed recovery.

the program returns the set of possibilities:

Action	¬aid	¬recovery
¬Action	¬aid	¬recovery
¬Action	aid	¬recovery
¬Action	aid	recovery

Hence, even though the medical aid would only have made it possible for the victim to recover, the accused's action nevertheless caused her death in that it prevented her recovery (see the first possibility). But if the denial of the aid was not a critical factor in the victim's death, then despite his heinous actions, the accused was convicted of a crime that he did not commit.

IX. GENERAL DISCUSSION

From Mill onwards, philosophers have argued that no logical distinction exists between causes and enabling conditions. The philosophical tradition has in turn led theorists to search for some other difference between enabling and causing. They have proposed many putative distinctions—enabling conditions are normal and causes abnormal, enabling conditions are common and causes rare, enabling conditions are constant and causes inconstant, and enabling conditions are explanatorily uninformative and causes explanatorily informative. These analyses entered legal theory with Hart and Honoré's influential study of causation and the law.[39] But, in fact, there is a genuine logical distinction between causes and enabling conditions in the *circumstances* of an observed state, that is, in the set of alternative possibilities that surround the state. Consider, for example, a dry match that is struck and lights. It is not necessary to strike it, because touching it with a lighted cigarette will do as well. The circumstances correspond to the following possibilities:

dry	struck	lights
dry	¬struck	lights
dry	¬struck	¬lights
¬dry	struck	¬lights
¬dry	¬struck	¬lights

Striking the match and its dryness are logically distinct here: Given that the match is dry, striking it causes it to light, but if it is not dry, it cannot be lit whether or not it is struck. Dryness is accordingly the enabling condition, and striking the match is the cause of it lighting. What determines causal status is the circumstances of the state, that is, the set of possibilities as a whole. The distinction between causing and allowing is, as this paper has shown, not always clear in the law. But when it is kept in mind, it can help to elucidate otherwise puzzling phenomena.

---

[39] *See id.*

Another misconception about causality is that the meaning of causal relations depends on a background generalization or on a framework of explanatory principles. In fact, causal relations are often justified by such principles, but the meanings of causal relations make no implicit reference to explanatory principles. You can therefore make a causal claim that denies their existence without self-contradiction:

> The water was turned into wine by a miracle that defies explanation.

However, the *inference* from correlation to causation is certainly strengthened by an account of an underlying mechanism or explanatory principle. The meaning of a causal relation is distinct from the reasons that justify its induction.

Theorists sometimes argue that causal relations cannot be deduced, but only induced[40]—a misconception that runs in parallel with the mistaken view that inferences about probabilities can only be induced.[41] However, a common sort of causal deduction occurs when background knowledge implies a causal interpretation of a state of affairs. You know that an insulin injection prevents a coma in diabetes, your diabetic friend gives herself such an injection, and you infer that the injection prevented her from going into a coma. Granted that individuals make causal deductions, theorists might suppose that they do so relying on laws of thought, that is, formal rules of inference of some sort. We have seen to the contrary that causal reasoning can be based on models. The illusory inferences we reported were predicted by the model theory, but cannot be explained by formal rules that support only valid deduction.

The morals of our results are twofold. On the one hand, they substantiate the model theory of causal relations, which is founded on straightforward assumptions that have been implemented in a computer program. On the other hand, they imply that theories of legal causation might be reformulated to be in

---

[40] *See* Cheng, *supra* note 28.

[41] *But see* P.N. Johnson-Laird et al., *Naive Probability: A Mental Model Theory of Extensional Reasoning*, 106 PSYCHOL. REV. 62 (1999).

closer concordance with how ordinary individuals think about causation. The advantage of such a foundation is that certain presently unclear legal notions of causation could be replaced by a few simple principles. The meanings of causal relations are sets of possibilities in which an effect cannot precede a cause. Naive individuals envisage these possibilities in mental models, and they make causal interpretations using their knowledge to envisage the circumstances of states of affairs. They infer the consequences of causal claims from their models of the premises.